

Crime as Prime

RICHARD HOLTON

According to a familiar picture, two elements are needed for guilt in most common law offences: the action, the *actus reus*, and the associated mental element, the *mens rea*. Neither is sufficient for the offence on its own. So a murder requires a killing—a homicide—and also an intention—*malice aforethought* as it is traditionally, and misleadingly, termed. And it is natural to think that these two elements can obtain independently. Thus there can be a killing that is not accompanied by the relevant intention—that is what happens in cases of accidental killing. Conversely, there can be an intention without a killing—that is what happens in attempted murder, and also in cases where so little is actually done that we do not even have an attempt.

Natural though this picture is, I think that we should reject it. I don't suggest that we should give up the notions of *actus reus* and *mens rea*. But I do suggest that we make the best sense of the law if we reject the idea that they are independent. The central idea I will develop here is that a crime like murder cannot be understood as the conjunction of a self-standing mental element and a self-standing physical element. Instead we should think of it as involving an intentional action—murder—which involves both elements inseparably intertwined.

My argument for this proceeds by way of an account of the puzzles that arise if we think of *mens rea* and *actus reus* as separable. One account concerns the notion just mentioned, that of attempt. Why is it—a much discussed question—that in many jurisdictions, attempt, even a completed attempt where the perpetrator thinks that they have done enough, is often punished less severely than success?¹ After all, if we have two independent

¹ See R.A. Duff, *Criminal Attempts* (Oxford: OUP 1997), ch. 4, for discussion both of how attempt is punished in various jurisdictions, and of possible justifications for it.

elements, the *mens rea* and the *actus reus*, it might seem, whether from considerations of natural justice or of deterrence, that it is the *mens rea* that matters; whether the act is successful or not is just down to luck. But if the *mens rea* is the same in each case, then the punishment should be the same. Yet we are, I think, naturally resistant to that conclusion. It isn't just that this is what the law happens to be. We think that this is what the law *should* be. We think that successful offences should be punished more severely than mere attempts; we are more inclined to morally condemn those who succeed than those who merely attempt; we feel more guilty about our own immoral successes than our mere attempts, and so on.

A second problem concerns the nature of *mens rea*. I said before that the traditional characterization of the *mens rea* for murder as *malice aforethought* is highly misleading: it has long been accepted that, as one authority puts it 'the "malice" may have in it nothing really malicious; and need never be really "aforethought"'.² Less misleadingly the mental element is standardly thought of as an intention. But that too is far from a perfect characterization. In many crimes a person can be guilty even though they lack what we would normally think of as an intention to commit the crime. So, to continue with the example of murder, if the death of the victim is sufficiently clearly foreseen by the accused, that may be sufficient to constitute the *mens rea*.³ Moreover, even if the accused does not expect the victim to die—if they would indeed be surprised by their death—an intention to inflict grievous bodily harm is sufficient for the *mens rea* for murder.⁴ And for many other crimes, though not in England for murder, recklessness is sufficient. How should we understand this? One possibility is that these various different criteria have no real unity: they have just been put together for pragmatic expediency. But the arguments that courts have used have typically been couched in terms of something like the development of the true nature of the crime, and of what it is, not to have an intention to murder, but to murder intentionally. This is reflected too in a common understanding

² C.S. Kenny, in J.C. Smith, *Smith and Hogan, Criminal Law*, 9th edition (London: Butterworths, 1999), p. 348.

³ Quite how likely it is foreseen as, and whether this should be a subjective or an objective standard, has been an issue of ongoing debate in the courts. Since *Woollin* [1989] AC82, [1998] Crim LR 890, the standard requirement in the UK is a subjective one of virtual certainty.

⁴ *Moloney* [1985] 1 AER 1025.

of what it is to do something intentionally. If you ask ordinary people whether the foreseen consequences of an agent's acts are things that they do intentionally, in many circumstances they will say that they are; a point to which we shall return.

A third problem arises when we put these two issues together. Although, as we have seen, the expectation that the victim will die is enough for the *mens rea* of murder, if the victim doesn't in fact die, it is not enough for the *mens rea* of attempted murder. (At least, not in England, and in most US jurisdictions; Scottish law is rather different, as is that of the State of Colorado.) Similarly, whilst intending to cause grievous bodily harm is sufficient for the *mens rea* for murder, attempting to cause grievous bodily harm is not sufficient for the *mens rea* for attempted murder.⁵ And whilst recklessness is sufficient for the *mens rea* of a crime like rape, it is not sufficient for attempted rape. Again, this might be thought to result from expediency. But again that doesn't seem quite right. Attempting to do something that one believes will result in a killing is not the same as intending to kill, and the law is right to treat them as different.⁶

So we are left with a puzzle. The idea I want to explore here is that we can make some progress on solving it by thinking of the relation between the *mens rea* and the *actus reus* in a very different way. Rather than conceiving of them as two independent elements which come together to constitute an offence, we should think of the offence as a single primitive notion, though one which entails (or presupposes) these two elements. Our way of understanding the *mens rea* involved in murder is as the mental state that is common to those who perpetrate it. The mental state of those guilty of attempted murder is a different thing, that of someone who unsuccessfully tries to kill.

⁵ *Smith and Horgan, Criminal Law*, p. 307. See R.A. Duff, *Intention, Agency, and Criminal Liability* (Oxford: Blackwells, 1990) and Duff, *Criminal Attempts* for illuminating discussion. In many ways the argument of this paper can be seen as an attempt to provide a theoretical underpinning for the position defended there. I hope that the main argument of this paper is also consistent with the position taken by Gideon Yaffe in his *Attempts* (Oxford: OUP, 2010), although my own view on why we should not distinguish tryings from attempts is somewhat different (see n.20 below). Yaffe's subsequent 'Criminal Attempts', *Yale Law Journal* 124 (2014): 92–156, also traces many of the difficulties in standard discussions on criminal attempt to the view that *mens rea* and *actus reus* can be identified independently; see especially pp. 118–20.

⁶ Though whether or not they should be punished in the same way, or indeed even subject to the same conviction, is a further question. See Yaffe, *Attempts*, pp. 119ff.

Such ideas may seem mysterious, but in suggesting this I am following a path that has become very familiar in many areas of philosophy. The main ideas have gone under various different names—‘externalism’, ‘disjunctivism’—and have been proposed with various different emphases. Here I want to take as my model the ‘Knowledge First’ account offered by Timothy Williamson.⁷ So I start by outlining that.

KNOWLEDGE

How should we conceive of knowledge? A traditional approach held it to be built out of the conjunction of three different elements: belief (something internal), truth (something external), and justification (either external or internal, depending on the account). But in a justly celebrated piece Gettier showed that this couldn’t be right.⁸ To use an example first offered by Russell: agents are surely justified in forming beliefs about the time by looking at a station clock. If the clock is telling the right time, they will have true justified beliefs. But suppose that the clock has, unprecedentedly, stopped, and someone glances at it exactly twelve hours after it stops, when it happens to be telling the right time. They will form a justified true belief that will not be knowledge.

Once we have the recipe other examples are easy to create. The initial response from philosophers was to try to give more conditions: knowledge is belief + truth + justification + something else. The outcome was not edifying. More complex accounts brought more complex counter-examples, until many began to doubt that an effective analysis would ever be given. Why should it? Analysis must stop somewhere. Why not think that the idea of knowledge, fundamental as it is, is primitive? It seems to be grasped by monkeys who cannot grasp the concept of belief.⁹ Why not think that it is one of our cognitive starting points?

Williamson’s *knowledge-first* account draws just such a conclusion. Knowledge should be taken as a starting point. Part of the justification for this is simple induction on the failure of attempted analyses. But that is a consideration in favour of thinking that knowledge is *unanalysable*.

⁷ Timothy Williamson, *Knowledge and Its Limits* (Oxford: Clarendon Press, 2000), chs. 1–3.

⁸ Edmund L. Gettier, ‘Is Justified True Belief Knowledge?’ *Analysis* 23 (1963): 121–3.

⁹ Drew Marticorena et al., ‘Monkeys Represent Others’ Knowledge but not their Beliefs’, *Developmental Science* 14 (2011): 1406–16. It used to be thought that the same was true of young children, but that is now in debate.

Here we are concerned with the rather different claim that it is *prime*. So let us first get clear on the difference between the two.

Neither notion is quite as straightforward as it might seem at first glance, and nor is their relation. Start with analysability. Take it as a two-place property and it's fairly clear: a notion is analysable iff it can be explained in some target vocabulary. That is what early proponents of the approach wanted to do: everything legitimate had to be analysed into the language of physics, or of sense-data or the like. But taken as a one-place property, and it is much less clear what constitutes a successful analysis. Everything can be analysed in terms of itself, so that is no good. There has to be some idea of a simpler vocabulary in which the analysis must be couched.

Now take primeness. Williamson introduced the term to this area, so he gets to say what it means. He defines prime conditions as those that are not composite; and composite conditions as those that consist of the conjunction of an internal condition with an external environmental condition. We can literally think of an internal condition as internal to the subject's skin. We could generalize the notion to say that a condition is composite relative to any two exclusive sets of conditions iff it is a conjunction of one drawn from each, but Williamson's initial characterization is good enough for us here.

Distinguish this account of primeness from the very different claim that prime concepts must provide the blocks from which all others are built. (Such a claim might perhaps be motivated by an over-literal parallel with the mathematical notion, and hence with the Fundamental Theorem of Arithmetic.) While it may be true that all analysable concepts are built out of unanalysable ones, there is no reason to think that prime concepts will be fundamental. I will make no such assumption.¹⁰

What is the relation between being prime and being unanalysable? It might seem that if something is not prime it will be analysable, since its decomposition will constitute an analysis. But that depends on whether the decomposition is sufficiently distinct from the target notion to count as one. Suppose that someone wanted to characterize being red in terms of reflecting a certain frequency of light and being judged to be red. Such a characterization is composite, but it is plausibly not an analysis

¹⁰ Note that, despite the 'knowledge first' slogan, and his gloss of belief in terms of knowledge, Williamson himself offers no such analytic claim: "a full-blown exact conceptual analysis of beliefs in terms of knows is too much to expect" (*Knowledge and Its Limits*, p. 47).

since the internal condition, being judged to be red, itself makes reference to the very notion of being red.¹¹ Conversely a condition can be analysable but not prime: it may be that the analysis does not split into the conjunction of an internal and an external condition (causal accounts of knowledge are like that: we shall return to them shortly). Primeness and unanalysability are independent notions.

Williamson offers a general argument for thinking that knowledge is not prime using a mix-and-match strategy. If knowledge really could be factored into these two elements, then it should be possible to take any two cases in which a certain thing is known, and come up with a new case of knowledge by combining the internal element from one and the external from the other. Yet this cannot in general be done. So the two elements are not independent.

Here is an adaptation of one example from the many that Williamson offers:

First Case. Suppose that Judy wants to know whether a party she missed was crowded. She gets testimony from two informants, A and B. A is trustworthy and Judy trusts him. He tells her that the party was indeed crowded and she believes him. B in contrast is utterly untrustworthy and wisely Judy puts no store in his testimony, but for reasons of his own in this case he truthfully tells her that the party was crowded (perhaps he is trying to make her regret having missed it). Since A is trustworthy, and is speaking the truth, and Judy accepts his testimony and comes to believe that the party was crowded, Judy gains knowledge that it was.

Second Case. This is just like the first, except that the characteristics of A and B are reversed, as are Judy's attitudes to them. This time it is B who is trustworthy, and it is B whom Judy trusts. Again Judy gains knowledge, but this time from B.

Third case. This is constructed from the first two cases by taking the features internal to Judy from the first, and the features external to her

¹¹ See Christopher Peacocke, 'Colour Concepts and Colour Experience', *Synthese* 66 (1984): 365–81 and Michael Smith, 'Peacocke on Red and Red', *Synthese* 68 (1986): 559–76 for the genuine debate here. There the account of redness under discussion is simply in terms of being judged to be red. That too is composite, but trivially so, since it consists of the conjunction of the internal state with any external state whatsoever; I gave a non-trivially composite version to avoid muddying the waters.

from the second. So Judy trusts A and distrusts B; yet it is B who is trustworthy and A who is not. Once again Judy forms the true belief that the party was crowded, but this time, since her belief is formed on the testimony of an utterly untrustworthy informant, it isn't knowledge.¹²

The Gettier cases showed us that in cases of knowledge there is a connection between what is known and the state of mind of the knower—intuitively this is what fails in cases like that of the station clock where the true belief is formed by chance. The mix-and-match argument shows us just how intimate that connection is: we cannot factor knowledge into independent internal and external elements. Whether we have a case of knowledge depends on the relations between those elements. Knowledge is *prime*.

Once we accept that knowledge is prime, we have to give up the idea that it is simply the conjunction of belief and something external. However, we do not have to deny that knowing entails belief. Indeed Williamson suggests that we can use the notion of knowledge to provide a rough gloss on belief: to believe a proposition is to treat it as if one knew it. But we cannot build a case of knowledge by adding truth to belief. (Can we turn a case of knowledge into one of false belief by subtracting truth? Yes, but only if that is understood temporally, not modally. If knowledge exists it can be lost because the world changes. But we cannot arrive at a case of knowledge by taking a case of false belief and then supposing that the thing believed had been true all along.)

MURDER

Now we turn from knowledge to criminal guilt. I suggest that the two are very similar. In particular, the relation between the *mens rea* and the *actus reus* in a crime like murder is much the same as that between the internal and the external elements in a case of knowledge. To see this, start by considering something like a Gettier case for murder:

D intends to kill V with poisonous tablets. The police have been watching; they remove the tablets and substitute some others that are apparently harmless. It turns out though that they contain a food colouring, to which V is, remarkably, and unknown to them, highly allergic. D gives them to V, who takes them and dies.

¹² Williamson, *Knowledge and Its Limits*, p. 72.

Is D intuitively guilty of murder? I can find no actual cases of quite this form.¹³ However, variations on such cases are standard in law exams, and those who mark them tell me that the expected answer is 'No'. The gap between D's intention and the actual killing is too great and too accidental; in this it resembles the case of the stopped station clock.¹⁴ D may be guilty of attempted murder, but not of murder.

Working backwards, we can construct from this a Williamson-style mix-and-match sequence:

First Case. D intends to kill with his own poisonous tablets; the police do not intervene; D gives V the tablets; V dies; D is guilty of murder. In this case, D is internally just like in the case we started with, but externally things are different.

Second Case. D intends to kill V; but he believes that the police are watching, and that they plan to substitute tablets. He also believes that V is allergic to the tablets they would substitute. Thinking this the perfect way to avoid liability for murder he allows things to take their course, and gives the substituted tablets to V. V dies. D is guilty of murder (though it would be doubtless hard to convict). In this case D is internally different to the case we started with, but externally things are the same.¹⁵

Third Case. The third case is just our original case, the same as the first case with respect to D's internal features, and the same as the second with respect to the features external to D. Yet here D is not guilty.

¹³ There are some cases that involve examples of more or less deviant causal chains. In *Thabo-Meli vs R* [1954] 1 WLR the victim died from exposure having been rolled off a cliff after the defendants thought they had killed him; there the Privy Council reasonably ruled that it would be 'much too refined a ground of judgement' to distinguish the initial assault from the subsequent treatment of the victim (thanks to John Gardner for the reference). *Nyuzi and Kudemera vs Republic (Malawi)* [1967] African Law Reports 249 involves a case (discussed in A. Kenny, *Freewill and Responsibility* (London: Routledge & Kegan Paul, 1978), pp. 12–21) in which a 'trial by ordeal' using an apparently harmless potion led to the death of four people who had been suspected of witchcraft; but there the charge, of which the accused was found guilty, was of witchcraft, not of murder, so the issue of the deviant causal chain was not germane to the finding.

¹⁴ The actual legal doctrine may be muddled by the doctrine of transferred intent (or transferred malice), although, since it is the same person involved, it should probably not be. Even if it were, I should be inclined to think that not so significant. In contrast to the other legal doctrines that we have been discussing, transferred intent is a doctrine that strikes many as artificial: Lord Mustill described it as 'useful enough to yield rough justice' while lacking 'any sound intellectual basis' [1998] 1 Cr App R 91 at 106.

¹⁵ Why not say that D *knew* about V's allergy, and the police plan? I don't want to assume that there could be a difference in knowledge without a difference in the external circumstances (though I don't think that Williamson's position requires this).

Is this so surprising? After all, it is a standard position, both in action theory and in the law, that causation matters: the mental state must cause the behaviour in the right sort of way. Are we simply seeing that mixing and matching disrupts the causal relationships, giving rise to deviant causal chains? I think that this is to understate the force of the challenge. The deviant causal chain approach encourages the thought that if we could understand the causal relation properly we could go back to a more complex conjunctive account. Murder = *mens rea* + *actus reus* + the right causal relation between them. The parallel attempt to reductively analyse the right causal relation has been made in both epistemology and in action theory, and in both it has failed dismally. No one has been able to specify the right kind of relation except by saying that it is whatever is needed to give rise to cases of knowledge or of action, and there is no reason to think that we will do any better in the case of murder. However, the problem with such an approach is not simply that it fails to provide a reductive account. It is that once one says that the relation is whatever is needed for knowledge, or for action, it is far from obvious that there will be a *causal* relation, even a non-reductively characterized one, that does the job (in cases of knowledge within abstract domains like mathematics, it is highly implausible that there is). While, in cases of action at least, there must be causation involved—Davidson showed us that—there is no reason to think that the causal relations constitute an independent third factor, rather than simply being further implications. Indeed it is unclear that we must accept that there is a non-reductive *relation* of any kind that fills a third place. All we can be sure of, in the case of knowledge, is that there is a belief, and the content of the belief is true, and that this is a case of knowledge; and in the case of action, that there is an appropriate mental state, and an appropriate bodily movement, and that this is a case of action. The primeness account provides an explanation of why we should be sceptical of the whole approach, of constructions using three conjuncts as well as of two.

My conclusion then is that, as with knowledge, so with murder: it is prime. It cannot be factorized into internal and external elements. But in some ways the conclusion is even more radical than it is for knowledge. As we saw, the claim that knowledge is prime is consistent with the idea that knowledge entails belief; and if, as Williamson suggests, to believe a proposition is to treat it as if one knew it, we have an explanation for the entailment (to know a proposition is certainly to treat it as if one knew it). So both cases of knowledge and of false belief can be seen to

have something—belief—in common. It seems unlikely that there is anything similar that can be said to unify cases of murder with cases of attempted murder. We cannot say in a parallel way that to attempt to murder someone is for it to be as if one actually murdered them. For attempts do not need to be completed attempts. And even apparently completed attempts will often differ from actual murders, since the latter will involve monitoring of how things are going, with a readiness to make further interventions if needed. Murder is an ongoing process. Even a frenzied attack typically takes time, involving different steps mediated by knowledge of what is happening—Is the victim escaping? Are they still moving? Is someone coming who can help them?—and finishing only when the victim is dead. An attempt to murder, even one that is in a sense complete, need not be like that.

Let me make one further clarification. In the Introduction to *Knowledge and Its Limits* Williamson pointed out a possible parallel between knowledge and desire; in subsequent work he has replaced talk of desire with talk of intention.¹⁶ The idea there is that one can parallel the relation between knowing that *p*, believing that *p*, and *p* being true, with the relation between performing the action of ϕ -ing, intending to ϕ , and one's ϕ -ing being successful. That is not what is being proposed here. Williamson's parallel is between two sets of schemas, those on the knowledge side substituting a sentence for '*p*', and those on the action side substituting a verb phrase for ' ϕ '. In contrast I am arguing for a parallel between the verb 'to know', and a variety of other action verbs, typified by 'to murder'. Mine is not a schematic proposal in the way that Williamson's is; it does not apply to each possible substitution for ' ϕ '. To see whether verbs have the similarities to 'to know' that I am interested in we need to examine them individually. Non-intentional verbs like 'to drop' and 'to kill' do not; we turn to this distinction next.

THE *MENS REA* FOR MURDER

My suggestion then is that in a case of murder the *mens rea* is essentially connected with the killing itself. It follows that the *mens rea* for murder, and the *mens rea* for attempted murder, are bound to be different. The issue

¹⁶ Timothy Williamson, 'Acting on Knowledge', in J.A. Carter, E. Gordon, and B. Jarvis (eds.), *Knowledge First: Approaches in Epistemology and Mind* (Oxford: OUP, 2017). Yair Levy has developed similar ideas: see 'Intentional Action First', *Australasian Journal of Philosophy* 91 (2013): 705–18.

is thus not why they are different; it is why they take the particular different forms that they do. Let's turn to that.

I have talked so far about criminal guilt. But the discussion can equally—perhaps preferably—be rephrased in terms of *action*. There are certain action verbs which we can think of as essentially intentional, in that they are only applicable if the agent is in a certain state of mind. Those expressing what have historically been the central offences of the criminal law are like this—to murder, to steal, to rape, to assault, to defraud—as are many others that have little place in the law—to borrow, to lie, to compliment. But other action verbs are not: one can kill someone or cause their death, drop something, wound someone, damage something, without needing to be in any particular mental state. Some offences make use of these verbs: causing death by dangerous driving for instance. And there are some offences—most obviously those involving negligence where it is distinguished from recklessness—that presuppose the *absence* of certain states of mind (though I cannot think of any ordinary verbs of action that do).

As a first thought, we can distinguish the intentional verbs from the others by saying that they can't readily be modified by the adverbs 'unintentionally' or 'inadvertently'. One can unintentionally kill someone or drop something; but there isn't a straightforward way in which one can unintentionally murder or steal. (Of course, if someone were to say something like that, we'd make some sense of what they said; but things would have to be stretched.) It is a plausible thesis that all intentional action verbs are, like 'murder' prime, though I shan't defend that here. I'll keep the focus on murder.

Understanding murder as an action verb, what might the *mens rea* for it be? I said at the outset that talk of an intention to kill is too restrictive. Recent philosophical discussions of intentions have identified them with plans: to have an intention to do something is to structure one's activities to try to ensure that it comes about, modifying those activities, and perhaps one's plans, if need be. That leaves space for a foreseen but unintended outcome, an outcome that one is not trying to bring about—one would not restructure one's activities to ensure that it did—but that one nonetheless believes will result from one's actions. But as we have seen, one can be guilty of murder if the death of the victim is a suitably closely foreseen consequence of one's action; and that is quite compatible with one not

having had an intention to kill. So the *mens rea* for murder cannot be the intention to kill. A fortiori, it cannot be the intention to murder.

Our earlier discussion of primeness should have shown that there is a general problem of understanding the *mens rea* as an intention. Intention is plausibly an internal state (at least, putting aside questions of the externality of referring terms and the like). But if murder cannot be factored into an internal and an external state, then in looking to intentions to characterize the *mens rea* we are looking at the wrong sort of thing.

Rather than trying to identify an independent mental state as the *mens rea*, we should focus on the action—and hence the crime—of murder. The *mens rea* just is whatever state of mind is entailed by that. As we refine our understanding of murder, so we refine our understanding of the *mens rea* involved. Of course most of the changes there have been driven by the courts; but then murder is a well-established notion, and most of the recent revisions have concerned relatively minor issues around the degree of foresight and the like. For an example that concerns much more general involvement, and rather less from the courts, consider sexual harassment. This is a relatively recent notion, dating to the early 1970s, and evolving considerably since then. What is the associated *mens rea*? That remains a matter of disagreement. Those who see it as a structural injustice tend to minimize the importance of the actor's attitudes, but even they are unlikely to discount them altogether: the difference between harassment and legitimate behaviour will have something to do with the state of mind of the actor.

The central idea here is of our moral and legal concepts evolving as we think more about them, and as circumstances change. We may be more or less objectivist about this process—may model it more or less closely on the process of discovering natural kinds in science. I will leave that issue unexplored. What is important for us is that the relevant *mens rea* and *actus reus* are driven by our conception of the crime, and not the other way around. We do not start with an independent conception of the wrongful intention and build the moral or legal concept from that.

To deny that the *mens rea* for an offence should be understood as an intention is not, however, to deny that we can characterize the action that constitutes the offence as intentional. Indeed, talk of intentional action brings the external dimension that talk of intentions lacks. Just as knowing something entails, or presupposes, that the thing known is

true, so intentionally performing some action obviously entails or pre-supposes that one performs that action. This is more promising.

If we switch to intentional actions, then we have two natural candidates for the *mens rea*: intentionally killing, and intentionally murdering. But if I am right that ‘murder’ is an intentional verb, then the second of these is something of a pleonasm. So let us focus instead on the first.

It might be thought that once again talk of intentionally killing is too restrictive: if one merely foresees that an action will result in a killing, the killing is not intentional. But in a much discussed set of experiments, Joshua Knobe found that most people don’t think that way, at least for closely related cases. Consider an executive who, motivated entirely by the goal of maximizing profit, embarks on a policy that he knows will also cause environmental damage. Does he intentionally harm the environment? Most people hold that he does. It seems likely then that they would say similar things about killing: that they would class a merely foreseen killing in such circumstances as intentional.¹⁷

Consider now the other half of Knobe’s findings. When asked to think about an otherwise identical case in which an executive embarks on a policy whose side effects are *beneficial* to the environment, most people hold that the executive does not intentionally help the environment. So how do we explain this? An obvious first thought is that bad foreseen consequences are counted as intentionally performed, whereas good foreseen consequences are not. However, another of Knobe’s experiments shows that that cannot be quite right either. A profit-driven executive whose actions have the side effect of violating a pernicious Nazi law—surely a good outcome in most subjects’ eyes—is judged to have acted intentionally; whereas a similarly driven executive whose actions have the bad effect of fulfilling the requirements of that law is not judged to have acted intentionally.¹⁸

So how should we explain these results? Elsewhere I have suggested that the asymmetry stems from a basic asymmetry in how we think about norms.¹⁹ On one side we are concerned with *violation* of a norm;

¹⁷ Joshua Knobe, ‘The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology’, *Philosophical Studies* 130 (2006): 203–31.

¹⁸ Joshua Knobe, ‘Reason Explanation in Folk Psychology’, *Midwest Studies in Philosophy* 31 (2007): 90–107.

¹⁹ Richard Holton, ‘Norms and the Knobe Effect’, *Analysis* 70 (2010): 417–24. For related ideas, see Philip Pettit, *The Robust Demands of the Good* (Oxford: OUP, 2015), ch. 6. For some

on the other with *following* it. When it comes to violation of a norm, our concern is with whether someone is prepared to knowingly violate it. Intentional violation amounts to just that: acting with disregard for the norm. In contrast, intentionally following a norm requires more. It requires that one's actions be guided by the norm. It is this asymmetry, I suggested, that explains Knobe's findings. It is sufficient for intentionally harming the environment that one intentionally—i.e. knowingly—violates a norm on harming it. But it is not sufficient for helping the environment that one merely acts in a way that conforms to a norm on helping it. To help it would require having one's behaviour guided by that norm, in a way that the executive's is not.²⁰

Let us return to murder, for now we have the pieces needed for a more convincing account. If the norm is not to kill, then intentional killing results from knowingly violating that norm. And one can intentionally violate it either by acting on a successful intention to kill, or by acting in such a way that the killing is foreseen. Arguably one also violates it by behaving with insufficient care: by behaving recklessly, or by doing something else—inflicting grievous bodily harm, for instance—that brings a real risk of death. The arguments in the courts that have surrounded these issues reflect, I suggest, our disagreement here.

Looked at in this way, the conditions on the *mens rea* for murder look as though they might be captured in the intuitive idea of intentionally killing. However, I'm still not sure that this is right. For obvious reasons, not every case of intentionally killing is a case of murder: not if it is done in self-defence, in legitimate warfare, as a result of provocation, and so on. Even if we insisted that in these cases the norm against killing is not violated, we surely cannot deny that they involve intentional killing. Nor does it help to say that since the *mens rea* is merely a necessary condition for the crime, it need not concern itself with the existence of defences. That doesn't help because these defences incorporate mental elements themselves. Thus self-defence requires that the accused honestly

plausible revisions to the account I offered, though in ways that I think are still compatible with what is said here, see B. Robinson, P. Stey, and M. Alfano 'Reversing the Side-Effect Effect: The Power of Salient Norms', *Philosophical Studies* 172 (2015): 177–206.

²⁰ One issue that I failed to address in the earlier paper is whether, in order to act with disregard for a norm, one needs to be aware of it. In cases like killing it is inconceivable that a sane agent would not. A genuine failure to recognize the norm against killing would be grounds for an insanity defence.

thinks the force used is both necessary and reasonable.²¹ The absence of such a belief is thus implicitly part of the *mens rea* for murder.

Perhaps then the best we can say is that the norm one violates is a norm against murder. Killings in legitimate warfare, in self-defence etc., don't look like violations of that. But then any attempt to understand murder in terms of something else has been lost. There is a general point here. It might be thought that any intentional verb can be analysed in terms of intentionally performing some non-intentional action. That thought could be maintained even given my earlier arguments that verbs like 'murder' are prime: it might be thought that the *mens rea* can't be characterized independently of the *actus reus*, but that the converse doesn't hold. So provided we can identify the *actus reus*, and then talk about performing it intentionally, we can give a non-composite, but nonetheless reductive, account of the crime. I suggest though that that thesis is false too. I see no way to analyse murder in general as the intentional performance of some action that can be characterized in non-intentional terms. And the same, I suggest, goes for other core criminal acts. Consider, for instance, how one would try to characterize the relevant unintentional action for theft: it is certainly not adequate to say that it is intentional permanent removal of the property of another without their consent (and even that involves intentional vocabulary in its talk of consent).

THE *MENS REA* FOR ATTEMPTED MURDER

Finally let us turn to the *mens rea* for attempted murder. I have argued that it cannot just be the *mens rea* for murder, since that cannot be disentangled from the *actus reus* of the killing. So we need to think of it as a self-standing thing. Now it would be possible to have an offence under which someone would be guilty if they either intended to kill, and acted on that intention, or if they intended to do some other thing, acted on that intention, and foresaw that a killing would result, or if they intended to inflict grievous bodily harm, and acted on that. But such an offence really would be a medley. In most jurisdictions, the legal doctrine of attempt is much more unified.²²

²¹ Williams (Gladstone) [1987] 78 Cr App R 276; Oatbridge [1984] 94 Cr App R 367.

²² Colorado's might be an exception. For discussion, see Yaffe, *Attempts*, pp. 47ff.

If we cannot understand attempt in terms of a successful crime minus success, we have to think instead of it as more self-standing. To get a grip on what it is we should start in the obvious place, with the ordinary idea of an attempt, that is, the ordinary idea of *trying* to do something.²³ Here I think we do need the idea of an intention, understood, as sketched above, as a plan. When one attempts to do something, the thing that one is attempting to do is what one has the intention to do. The intention brings a commitment to a certain course of action; the attempt is an attempt to fulfil that commitment.²⁴ But then we can immediately see why mere foresight does not provide the *mens rea* for attempted murder, since mere foresight, whilst it might be enough for doing something intentionally, is not enough for an intention; one is not committed to bringing about what one merely foresees.²⁵ Similarly, intentionally causing grievous bodily harm is not enough for an intention to kill, so attempting to cause grievous bodily harm is not enough for attempted murder.

There is of course more to say. The mere possession of an intention is not enough, for the idea of attempt is essentially action involving: one does not have an attempt until the plan has been put into motion. And I suspect that once again we have the structure with which we have become familiar: attempt entails an intention, and it entails action, but it cannot be factored into those two elements, since each is too intimately involved with the other. We have, not just intention, but intention-in-action. Quite how far down the path of action someone needs to go

²³ See Yaffe, *Attempts*, ch. 3, and 'Criminal Attempts' for very helpful applications of the notion of trying to the doctrine of attempt. Note that Michael Bratman, *Intention, Plans and Practical Reason* (Cambridge, MA: Harvard University Press, 1987), pp. 113ff. denies that trying to do something entails intending to do it. Like Yaffe I find that implausible. My own reason for doing so is rather different to Yaffe's: I deny that forming inconsistent intentions has to be irrational. See my *Willing, Wanting, Waiting* (Oxford: Clarendon Press, 2009), ch. 2.

²⁴ See Yaffe, 'Criminal Attempts' for a sophisticated development of this idea. Yaffe there divides the relevant commitment into a commitment to promote, a commitment not to reconsider, and a commitment not to feel regret, which can hold independently. I'm not altogether convinced of the need for such sophistication—I suspect that even in the law we can get away with what Yaffe calls the 'narrow sense' of intention—but I shan't argue for this here.

²⁵ For the classic articulation of the distinction, see Bratman, *Intention, Plans and Practical Reason*, ch. 8. Note that while in the cases discussed above Knobe and others have found that ordinary subjects are prepared to say that agents who pollute with foresight pollute intentionally, they are much less ready to say that they have intentions to pollute; see H. McCann, 'Intentional Action and Intending: Recent Empirical Studies', *Philosophical Psychology* 18 (2005): 737–48.

before mere preparation turns into enough of an attempt to be actionable is doubtless vague; but it certainly requires more than the possession of the intention alone.²⁶

Let me finish this section by asking why this matters. Suppose it is true that the idea of attempt in most jurisdictions is more unified than the medley offence that I described several paragraphs back, which tries to understand it in terms of the *mens rea* that would be present for the successful crime. What significance should we assign to that? What reason do we have to think that the law works with the more unified notions, and hence that an account in terms of such notions is more likely to be accurate? And more pressing still, even if it does, why should it?

Obviously these are big questions. My answer is that insofar as the law is genuinely trying for an offence of attempt, as we ordinarily understand it, this is what it needs to do. I have argued that the common law works with something like the moral equivalent of natural kinds. This is so for the substantial offences—murder, rape, theft, and so on, and it is equally true of the idea of attempt. We can work to understand those ideas, and to see how they fit together, which have priority, and so on. Sometimes we discover new ones (I have mentioned sexual harassment), and sometimes we come to think that those we have been using are mistaken (the idea of honour thankfully has a smaller role in our current thinking than it once did). But if the development of moral philosophy has shown us anything, it has shown us that attempts to throw out these categories and start afresh are hopeless. In this though, the law is no different to other kinds of thought. We are all in Neurath's boat.²⁷

²⁶ Again, see Yaffe, 'Criminal Attempts', pp. 116–18 for discussion, and for a clear argument that to state an intention is not, in the relevant sense, to act on it.

²⁷ This paper was presented at the Human Rights and the Human Mind conference in Tel Aviv, January 2014; thanks to Amit Pundik, my commentator there, and the audience members. An early version was published in the proceedings of that conference in *Law and Ethics of Human Rights* 9 (2015): 181–93. The material was subsequently presented at the Intention workshop at Leeds in February 2014, at the Criminal Law's Person conference at York in December 2015, and at Yale in March 2017; thanks to Yair Levy, my commentator at Leeds, and the audience members on all those occasions. Special thanks to Tim Williamson for discussion and comments on the written version, Gideon Yaffe, Michael Moore, and the referees for *LEHR*.